WHAT IS CLAIMED IS:

1. A method for at least one of genotyping and haplotyping a sequence of polymorphic genetic loci in a deoxyribonucleic acid (DNA) sample or identifying a strain variant from the DNA sample, comprising:
   i) providing one or more microarrays that include a set of oligonucleotide probes that are capable of detecting the at least one of the genotypes and the haplotypes or the strain variant;
   ii) hybridizing the DNA sample to the one or more microarrays to create a hybridization pattern; and
   iii) determining at least one of a genotype and a haplotype or a strain variant based on the hybridization pattern.

2. The method of claim 1, wherein the one or more microarrays include a set of oligonucleotide probes that are capable of detecting at least one of all known genotypes and all known haplotypes at the polymorphic genetic loci or the strain identification.

3. The method of claim 1, wherein the one or more microarrays are configured to include at least one of an optimal set and an optimal arrangement of oligonucleotide probes.

4. The method of claim 3, further comprising optimizing the least one of the set and arrangement of oligonucleotides based on the following:

$$\min \sum_{type\ j} w_j \ E\left[\Pi_{T_j \neq \hat{T}_j}\right]$$
$$\Leftrightarrow \ \min \sum_{type\ j} w_j \ \Pr\left(T_j \neq \hat{T}_j\right).$$

wherein T_j is a true allele contained in the DNA sample, $\hat{T}_j$ is the allele determined by the hybridization step, $\Pi_x = \begin{Bmatrix} 1, & \text{if } X \text{ is true} \\ 0, & \text{otherwise} \end{Bmatrix}$, and w_j is a weight assigned to at least one of the genotype and the haplotype j.

5. The method of claim 4, wherein the weights are provided as follows: $w_j = 1 \forall_j$, wherein $\forall_j$ is a set of at least one of all known genotypes and all known haplotypes at one or more predetermined polymorphic genetic loci.

6. The method of claim 4, wherein the weights are provided as follows: w_j is different for each genotype or haplotype.

7. The method of claim 4, wherein step (iii) produces a vector of n measurements, wherein n is a number of probes contained on the one or more microarrays.

8. The method of claim 7, wherein the n potential probes provided to identify N known genotypes or haplotypes are each associated with a response vector $\vec{v}_j \in \{0,1\}^N, j = 1,\ldots,n$.

9. The method of claim 8, further comprising generating a graph G on vertices corresponding to probe response vectors.

10. The method of claim 9, wherein the graph G is a complete edge-weighted and vertex-weighted undirected graph $G = (V, E)$ provided on n vertices, wherein n is the number of potential probes.

11. The method of claim 10, wherein the weights w of each vertex v and each edge e are constrained by: $0 \leq w(v), w(e) \leq 1$.

12. The method of claim 11, wherein the weight w of a vertex v is set to:

$$w(v) = \min\{\text{fraction of 0's, fraction of 1's}\}.$$

13. The method of claim 11, wherein the weight w of an edge $e = \{u,v\}$ is set to:

$$w(e) = \text{Hamming distance/vector length},$$

wherein Hamming distance is measured between the probe response vectors corresponding to vertices u and v, and vector length is the length of said probe response vectors, namely, N.

14. The method of claim 10, further comprising modifying the graph G by thresholding the edges such that the modified graph G_{mod} is defined as $G_{mod} =$

(V, E_{mod}), wherein $E_{mod} = \{e \in E: w(e) \leq \rho\}$, and ρ is a selected threshold value.

15. The method of claim 14, wherein, for the modified graph G_{mod} and the probe set size M, the following is performed:

    i) initializing a current-best list of independent sets with associated information weights,

    ii) initializing vertex boosting weights to vertex weights $w(v)$,

    iii) defining a probability distribution on the vertex subset based on vertex boosting weights,

    iv) choosing a random subset of vertices of a specified size M based on the probability distribution,

    v) eliminating one of the end-point vertices in each of the edges remaining in the induced subgraph on the random subset,

    vi) modifying the vertex boosting weights by increasing the weights of the vertices that are retained in the subset and decreasing the weights of the vertices that were selected in step (iv) but eliminated in step (v), and

    vii) repeating steps (iii) through (vi) for at least one of a predetermined number of iterations and until no improvement to the list of top independent sets is achieved.

16. The method of claim 15, wherein, for the modified graph G_{mod} and the probe set size M, steps (ii) through (vii) are repeated for a predetermined number of iterations, each iteration starting with reinitializing vertex boosting weights to vertex weights $w(v)$ in step (ii).

17. The method of claim 16, wherein, for a given fixed small $0<\in<<1$, the probe set size M satisfies an inequality $Pr(\forall$code pairs, Hamming distance $\geq 1) > 1-\in$.

18. The method of claim 16, wherein, for a given fixed small $0<\epsilon<<1$ and a fixed $\alpha>1$, the probe set size M satisfies an inequality $Pr(\forall$code pairs, Hamming distance $\geq \alpha) > 1-\epsilon$.

19. The method of claim 15, wherein the threshold ρ has a value to enable the graph G to have a sparsity bounded by $A \leq$ sparsity $\leq B$, wherein the sparsity is definable by the average degree of a vertex in the graph G.

20. The method of claim 19, wherein the lower bound A is a relatively small constant, and the upper bound B is a function of the number of vertices n.

21. A software arrangement which, when executed on a processing device, configures the processing device to perform the steps comprising:

i) hybridizing the DNA sample to one or more microarrays to create a hybridization pattern, the one or more microarrays including a set of oligonucleotide probes that are capable of detecting at least one set of genotypes and haplotypes for a sequence of polymorphic genetic loci in a deoxyribonucleic acid (DNA) sample or identifying a strain variant from the DNA sample; and

ii) determining at least one of a genotype and a haplotype or a strain variant based on the hybridization pattern.

22. The software arrangement of claim 21, wherein the one or more microarrays include a set of oligonucleotide probes that are capable of detecting at least one of all known genotypes and all known haplotypes at the polymorphic genetic loci or strain variation.

23. The software arrangement of claim 21, wherein the one or more microarrays are configured to include at least one of an optimal set and an optimal arrangement of oligonucleotide probes.

24. The software arrangement of claim 23, wherein the processing device is further configured to optimize the least one of the set and arrangement of oligonucleotides based on the following:

$$\min \sum_{type\ j} w_j \ E\left[\Pi_{T_j \neq \hat{T}_j}\right]$$
$$\Leftrightarrow \ \min \sum_{type\ j} w_j \ \Pr\left(T_j \neq \hat{T}_j\right).$$

wherein T_j is a true allele contained in the DNA sample, $\hat{T}_j$ is the allele determined by the hybridization step, $\Pi_x = \begin{Bmatrix} 1, & \text{if } X \text{ is true} \\ 0, & \text{otherwise} \end{Bmatrix}$, and w_j is a weight assigned to at least one of the genotype and the haplotype j.

25. The software arrangement of claim 24, wherein the weights are provided as follows: $w_j = 1 \forall_j$, wherein $\forall_j$ is a set of at least one of all known genotypes and all known haplotypes at one or more predetermined polymorphic genetic loci.

26. The software arrangement of claim 24, wherein the weights are provided as follows: w_j is different for each genotype or haplotype.

27. The software arrangement of claim 26, wherein step (i) produces a vector of n measurements, wherein n is a number of probes contained on the microarray.

28. The software arrangement of claim 26, wherein the n potential probes provided to identify N known genotypes or haplotypes are each associated with a response vector $\vec{v}_j \in \{0,1\}^N$, $j = 1,\ldots,n$.

29. The software arrangement of claim 28, wherein the processing device is further configured to generate a graph G on vertices corresponding to probe response vectors.

30. The software arrangement of claim 29, wherein the graph G is a complete edge-weighted and vertex-weighted undirected graph $G = (V, E)$ provided on n vertices, wherein n is the number of potential probes.

31. The software arrangement of claim 30, wherein the weights w of each vertex v and each edge e are constrained by: $0 \leq w(v), \ w(e) \leq 1$.

32. The software arrangement of claim 31, wherein the weight w of a vertex v is set to:

$$w(v) = \min\{\text{fraction of 0's, fraction of 1's}\}/100.$$

33. The software arrangement of claim 31, wherein the weight w of an edge $e = \{u,v\}$ is set to:

$$w(e) = \text{Hamming distance/vector length},$$

wherein Hamming distance is measured between the probe response vectors corresponding to vertices u and v, and vector length is the length of said probe response vectors, namely, N.

34. The software arrangement of claim 30, wherein the processing device is further configured to modify the graph G by thresholding the edges such that the modified graph G_{mod} is defined as $G_{mod} = (V, E_{mod})$, wherein $E_{mod} = \{e \in E\colon w(e) \le \rho\}$, and ρ is a selected threshold value.

35. The software arrangement of claim 34, wherein, for the modified graph G_{mod} and the probe set size M, the following is performed:

i) initializing a current-best list of independent sets with associated information weights,

ii) initializing vertex boosting weights to vertex weights $w(v)$,

iii) defining a probability distribution on the vertex subset based on vertex boosting weights,

iv) choosing a random subset of vertices of a specified size M based on the probability distribution,

v) eliminating one of the end-point vertices in each of the edges remaining in the induced subgraph on the random subset,

vi) modifying the vertex boosting weights by increasing the weights of the vertices that are retained in the subset and decreasing the weights of the vertices that were selected in step (iv) but eliminated in step (v), and

vii) repeating steps (iii) through (vi) for at least one of a predetermined number of iterations and until no improvement to the list of top independent sets is achieved.

36. The software arrangement of claim 35, wherein, for the modified graph G_{mod} and the probe set size M, steps (ii) through (vii) are repeated for a predetermined number of iterations, each iteration starting with reinitializing vertex boosting weights to vertex weights $w(v)$ in step (ii).

37. The software arrangement of claim 36, wherein, for a given fixed small $0<\epsilon<<1$, the probe set size M satisfies an inequality $Pr(\forall$code pairs, Hamming distance $\geq 1) > 1-\epsilon$.

38. The software arrangement of claim 36, wherein, for a given fixed small $0<\epsilon<<1$ and a fixed $\alpha>1$, the probe set size M satisfies an inequality $Pr(\forall$code pairs, Hamming distance $\geq \alpha) > 1-\epsilon$.

39. The software arrangement of claim 36, wherein the threshold ρ has a value to enable the graph G to have a sparsity bounded by $A \leq$ sparsity $\leq B$, wherein the sparsity is definable by the average degree of a vertex in the graph G.

40. The software arrangement of claim 39, wherein the lower bound A is a relatively small constant, and the upper bound B is a function of the number of vertices n.

41. A storage medium which includes thereon a software arrangement for providing one or more microarrays, which is capable of configuring a processing arrangement to perform the steps comprising:

i) receiving information regarding a hybridization of the DNA sample to one or more microarrays to create a hybridization pattern, the one or more microarrays including a set of oligonucleotide probes that are capable of detecting at least one set of genotypes and haplotypes for a sequence of polymorphic genetic loci in a deoxyribonucleic acid (DNA) sample or identifying a strain variant from the DNA sample; and

ii) determining at least one of a genotype and a haplotype or a strain variant based on the hybridization pattern.

42. The storage medium of claim 41, wherein the one or more microarrays include a set of oligonucleotide probes that are capable of detecting at least one of all known genotypes and all known haplotypes at the polymorphic genetic loci or strain variation.

43. The storage medium of claim 42, wherein the one or more microarrays are configured to include at least one of an optimal set and an optimal arrangement of oligonucleotide probes.

44. The storage medium of claim 43, wherein the processing device is further configured to optimize the least one of the set and arrangement of oligonucleotides based on the following:

$$\min \sum_{type\, j} w_j \; E\left[\Pi_{T_j \neq \hat{T}_j}\right]$$
$$\Leftrightarrow \; \min \sum_{type\, j} w_j \; \Pr\left(T_j \neq \hat{T}_j\right).$$

wherein T_j is a true allele contained in the DNA sample, $\hat{T}_j$ is the allele determined by the hybridization step, $\Pi_X = \begin{Bmatrix} 1, & \text{if } X \text{ is true} \\ 0, & \text{otherwise} \end{Bmatrix}$, and w_j is a weight assigned to at least one of the genotype and the haplotype j.

45. The storage medium of claim 44, wherein the weights are provided as follows: $w_j = 1 \forall_j$, wherein $\forall_j$ is a set of at least one of all known genotypes and all known haplotypes at one or more predetermined polymorphic genetic loci.

46. The storage medium of claim 44, wherein the weights are provided as follows: w_j is different for each genotype or haplotype.

47. The storage medium of claim 44, wherein step (i) produces a vector of n measurements, wherein n is a number of probes contained on the microarray.

48. The storage medium of claim 46, wherein the n potential probes provided to identify N known genotypes or haplotypes are each associated with a response vector $\vec{\upsilon}_j \in \{0,1\}^N$, $j = 1,\ldots,n$.

49. The storage medium of claim 48, wherein the processing device is further configured to generate a graph G on vertices corresponding to probe response vectors.

50. The storage medium of claim 49, wherein the graph G is a complete edge-weighted and vertex-weighted undirected graph $G = (V, E)$ provided on n vertices, wherein n is the number of potential probes.

51. The storage medium of claim 50, wherein the weights w of each vertex v and each edge e are constrained by: $0 \leq w(v), w(e) \leq 1$.

52. The storage medium of claim 51, wherein the weight w of a vertex v is set to:

$$w(v) = \min\{\text{fraction of 0's, fraction of 1's}\}.$$

53. The storage medium of claim 52, wherein the weight w of an edge $e = \{u,v\}$ is set to:

$$w(e) = \text{Hamming distance/vector length},$$

wherein Hamming distance is measured between the probe response vectors corresponding to vertices u and v, and vector length is the length of said probe response vectors, namely, N.

54. The storage medium of claim 53, wherein the processing device is further configured to modify the graph G by thresholding the edges such that the modified graph G_{mod} is defined as $G_{mod} = (V, E_{mod})$, wherein $E_{mod} = \{e \in E: w(e) \leq \rho\}$, and ρ is a selected threshold value.

55. The storage medium of claim 54, wherein, for the modified graph G_{mod} and the probe set size M, the following is performed:

i) initializing a current-best list of independent sets with associated information weights,

ii) initializing vertex boosting weights to vertex weights $w(v)$,

iii) defining a probability distribution on the vertex subset based on vertex boosting weights,

iv) choosing a random subset of vertices of a specified size M based on the probability distribution,

v) eliminating one of the end-point vertices in each of the edges remaining in the induced subgraph on the random subset,

vi) modifying the vertex boosting weights by increasing the weights of the vertices that are retained in the subset and decreasing the weights of the vertices that were selected in step (iv) but eliminated in step (v), and

vii) repeating steps (iii) through (vi) for at least one of a predetermined number of iterations and until no improvement to the list of top independent sets is achieved.

56. The storage medium of claim 54, wherein, for the modified graph G_{mod} and the probe set size M, steps (ii) through (vii) are repeated for a predetermined number of iterations, each iteration starting with reinitializing vertex boosting weights to vertex weights $w(v)$ in step (ii).

57. The storage medium of claim 54, wherein, for a given fixed small $0<\epsilon<<1$, the probe set size M satisfies an inequality $Pr(\forall$code pairs, Hamming distance $\geq 1) > 1-\epsilon$.

58. The storage medium of claim 54, wherein, for a given fixed small $0<\epsilon<<1$ and a fixed $\alpha>1$, the probe set size M satisfies an inequality $Pr(\forall$code pairs, Hamming distance $\geq \alpha) > 1-\epsilon$.

59. The storage medium of claim 54, wherein the threshold ρ has a value to enable the graph G to have a sparsity bounded by $A \leq$ sparsity $\leq B$, wherein the sparsity is definable by the average degree of a vertex in the graph G.

60. The storage medium of claim 59, wherein the lower bound A is a relatively small constant, and the upper bound B is a function of the number of vertices n.

61. A system for at least one of genotyping and haplotyping polymorphic genetic loci or strain identification in a deoxyribonucleic acid (DNA) sample, comprising:

a processing arrangement which is capable of being programmed to:

i) receive information regarding a hybridization of the DNA sample to one or more microarrays to create a hybridization pattern, the one or more microarrays including a set of oligonucleotide probes that are capable of detecting at least one set of genotypes and haplotypes for a sequence of polymorphic genetic loci in a deoxyribonucleic acid (DNA) sample or identifying a strain variant from the DNA sample; and

ii) determine at least one of a genotype and a haplotype or a strain variant based on the hybridization pattern.

62. The system of claim 61, wherein the one or more microarrays include a set of oligonucleotide probes that are capable of detecting at least one of all known genotypes and all known haplotypes at the polymorphic genetic loci or strain variation.

63. The system of claim 61, wherein the one or more microarrays are configured to include at least one of an optimal set and an optimal arrangement of oligonucleotide probes.

64. The system of claim 63, wherein the processing arrangement is further programmed to optimize the least one of the set and arrangement of oligonucleotides based on the following:

$$\begin{aligned} & \min \sum_{type\ j} w_j \; E\left[\Pi_{T_j \neq \hat{T}_j}\right] \\ \Leftrightarrow \; & \min \sum_{type\ j} w_j \; \Pr\left(T_j \neq \hat{T}_j\right). \end{aligned}$$

wherein T_j is a true allele contained in the DNA sample, $\hat{T}_j$ is the allele determined by the hybridization step, $\Pi_X = \begin{Bmatrix} 1, & \text{if } X \text{ is true} \\ 0, & \text{otherwise} \end{Bmatrix}$, and w_j is a weight assigned to at least one of the genotype and the haplotype j.

65. The system of claim 64, wherein the weights are provided as follows: $w_j = 1 \forall_j$, wherein $\forall_j$ is a set of at least one of all known genotypes and all known haplotypes at one or more predetermined polymorphic genetic loci.

66. The system of claim 64, wherein the weights are provided as follows: w_j is different for each genotype or haplotype.

67. The system of claim 64, wherein step (i) produces a vector of n measurements, wherein n is a number of probes contained on the one or more microarrays.

68. The system of claim 66, wherein the n potential probes provided to identify N known genotypes or haplotypes are each associated with a response vector $\vec{v}_j \in \{0,1\}^N, j = 1,...,n$.

69. The system of claim 68, wherein the processing arrangement is further programmed to generate a graph G on vertices corresponding to probe response vectors.

70. The system of claim 69, wherein the graph G is a complete edge-weighted and vertex-weighted undirected graph $G = (V, E)$ provided on n vertices, wherein n is the number of potential probes.

71. The system of claim 70, wherein the weights w of each vertex v and each edge e are constrained by: $0 \leq w(v), w(e) \leq 1$.

72. The system of claim 71, wherein the weight w of a vertex v is set to:

$$w(v) = \min\{\text{fraction of 0's, fraction of 1's}\}.$$

73. The system of claim 71, wherein the weight w of an edge $e = \{u,v\}$ is set to:

$$w(e) = \text{Hamming distance/vector length},$$

wherein Hamming distance is measured between the probe response vectors corresponding to vertices u and v, and vector length is the length of said probe response vectors, namely, N.

74. The system of claim 70, wherein the processing arrangement is further programmed to modify the graph G by thresholding the edges such that the modified graph G_{mod} is defined as $G_{mod} = (V, E_{mod})$, wherein $E_{mod} = \{e \in E\text{: } w(e) \leq \rho\}$, and ρ is a selected threshold value.

75. The system of claim 74, wherein, for the modified graph G_{mod} and the probe set size M, the following is performed:

i) initializing a current-best list of independent sets with associated information weights,

ii) initializing vertex boosting weights to vertex weights $w(v)$,

iii) defining a probability distribution on the vertex subset based on vertex boosting weights,

iv) choosing a random subset of vertices of a specified size M based on the probability distribution,

v) eliminating one of the end-point vertices in each of the edges remaining in the induced subgraph on the random subset,

vi) modifying the vertex boosting weights by increasing the weights of the vertices that are retained in the subset and decreasing the weights of the vertices that were selected in step (iv) but eliminated in step (v), and

vii) repeating steps (iii) through (vi) for at least one of a predetermined number of iterations and until no improvement to the list of top independent sets is achieved.

76. The system of claim 75, wherein, for the modified graph G_{mod} and the probe set size M, steps (ii) through (vii) are repeated for a predetermined number of iterations, each iteration starting with reinitializing vertex boosting weights to vertex weights $w(v)$ in step (ii).

77. The system of claim 76, wherein, for a given fixed small $0<\epsilon<<1$, the probe set size M satisfies an inequality $Pr(\forall$code pairs, Hamming distance $\geq 1) > 1-\epsilon$.

78. The system of claim 76, wherein, for a given fixed small $0<\epsilon<<1$ and a fixed $\alpha>1$, the probe set size M satisfies an inequality $Pr(\forall$code pairs, Hamming distance $\geq \alpha) > 1-\epsilon$.

79. The system of claim 76, wherein the threshold ρ has a value to enable the graph G to have a sparsity bounded by $A \leq$ sparsity $\leq B$, wherein the sparsity is definable by the average degree of a vertex in the graph G.

80. The system of claim 79, wherein the lower bound A is a relatively small constant, and the upper bound B is a function of the number of vertices n.